Prokaryotic Transcription Regulation

Introduction

Most sources teach this before eukaryotic gene regulation. I choose to put it after, because it gets SO specific that it's easy to lose sight of the forest for the trees. You have to know both. This is actually pretty simple, but since bacteria are so easy to study, we studied the crap out of this process. Which means we know a lot about it, which means we like to ask questions about it.

Lactose Operon

The players on this stage are the **operator** (DNA sequence within transcription unit), the **promoter** (where RNA polymerase binds), the **lac-i** gene that makes the **inhibitory-protein** that binds the operator (the stop signal), and the **CAP protein** (the go signal).

You are a bacterium. You like glucose. Nom nom nom. Glucose. Energy. You don't like lactose. Boo lactose. But, you're no fool, you'd rather eat lactose than starve, so you will tolerate lactose, but only if you REALLY have to. So if you usually have glucose, and you don't even like lactose, you probably wouldn't dedicate energy to making the stuff that helps you eat lactose. You certainly want it in your back pocket, ready to turn on when you need it, but most of the time, it would be a waste. So, by default, you say no lactose, bad lactose, and lactose machinery is silenced.

So when would you want that machinery on? When you have no glucose AND there is lactose. Enter the lac operon. Man, can this get complicated, but it doesn't have to.

There is a **polycistronic** DNA operon, all under the control of one promoter region. Typical polycistronic DNA. And if that operon is on, you make all the things you need to eat and digest the lactose. The coding sequence has a promoter region where RNA polymerase binds, with a TATA box, just like everywhere. But it's what OTHER STUFF does to that operon that messes with RNA polymerase that makes this system so interesting.

The lac operon consists of an **upstream promoter** (activating) and a **downstream operator** (inhibitory), around the +1 site. The actual genes of the operon start past the downstream operator.

Your default is off (you don't like lactose, you don't see any anyway). This is achieved by turning the lac operon off. The way to keep it off by default (as in all the time) is with an **inhibitory protein** made by an inhibitory gene that is **constitutively on**. Constitutively on means always on and can't be turned off. This is the lac-i gene, or the i-protein, (i is for inhibitory). The lac-i gene is WAY upstream from the lac operon. The product of that gene, the i-protein acts within the lac operon at the operator. The **inhibitory protein** binds to the **operator region DOWNSTREAM of the promoter region**. No matter what, if the lac-i gene is on, and the i-protein is made, there will be a physical barrier preventing RNA polymerase from going forward. So to ever even consider RNA polymerase coding that operon, the brake has to be removed. To do that, all that i-protein has to be bound up. The lac-i gene can't be turned off; it's constitutively on. So instead, all the i-protein has to be sequestered so it can't get to the operator. That happens when **lactose is abundant**. If there's a lot of lactose around, the lactose binds up the i-protein inhibitor, and nothing binds the operator. If lactose is scarce, i-protein is not bound up, it binds the operator, and the operon is inhibited.

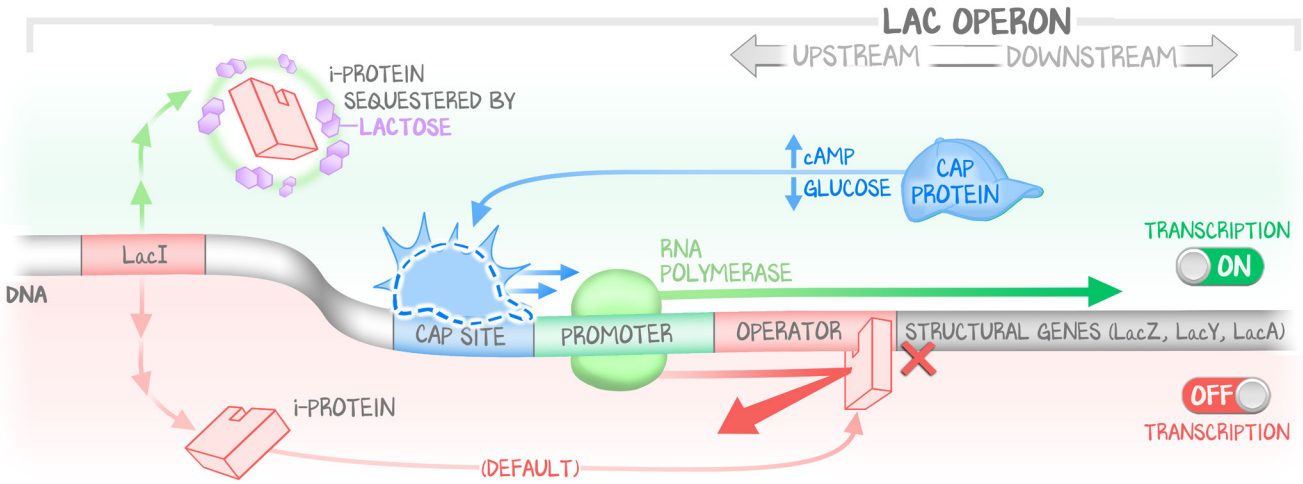

Figure 11.1: Lac Operon
Lac operon demonstrating lac-i gene distal to the operon, the i-protein binding to the operator between the promoter and the genes of the operon, and the CAP protein, binding upstream of the promoter to initiate transcription.

But remember, just because lactose is there doesn't mean you want to eat it. You don't like lactose. You like glucose. So even if there is plenty of lactose, as long as there's just a little bit of glucose, you'd rather have that, so your lac operon stays off. Even if high lactose levels remove the stop from the operator region, there is no go signal for RNA polymerase; nothing is bound to the promoter region.

To get the lac operon going, the cell will also need to run out of glucose. The way the cell recognizes that is by the levels of cAMP. When glucose is present, ATP is made. When no glucose and more ATP can't be made, and as ATP is used, cAMP levels rise. A special protein called the **CAP protein** is normally off. But when cAMP levels rise, it turns on. CAP binds the promoter.

And at that point, with both no glucose so cAMP rises activating the CAP-protein AND there's lots of lactose around so the i-protein is sequestered, then the gene is transcribed. For gene transcription, the **CAP protein** must be bound to the promoter (low glucose, high cAMP), and **i-protein must NOT** be bound to the operator (lots of lactose).

The promoter-region-needing-a-transcription-factor isn't new. What's special about this system is that **i-protein is not an effector**; it doesn't influence the RNA polymerase, it's like a concrete wall that needs to be removed.

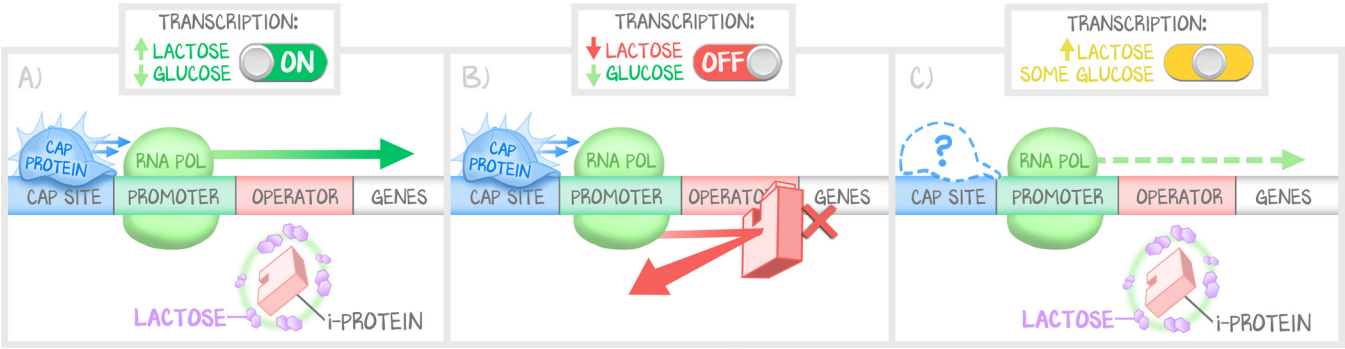

Figure 11.2: Lac Operon in Various States
(1) If the CAP protein is on and the operator not bound to i-protein, the operon is transcribed. (2) Regardless of the state of the CAP protein, if the i-protein is bound to the operator, there will be no expression of the gene. (3) If the i-protein is not bound to the operator, a small amount of the operon is transcribed, but not enough to matter, if CAP protein is off.

The whole point here is to show that there is a promoter region (like an enhancer in eukaryotes) upstream of the promoter region. There is also an operator region (unique to prokaryotes) after the promoter region. These two mechanisms are regulated by intracellular levels of stuff (CAP cAMP, inhibitor lactose) and it's the interplay between inhibitor/suppressor and promoter/enhancer that ultimately decides whether the operon is transcribed or not. Again, what's special about this system is that **i-protein is not an effector**; it doesn't influence the RNA polymerase, it is like a concrete wall that needs to be removed.

Trp Operon

The trp operon is another example of operators inhibiting transcription, but with a little twist. The trp repressor is made by a gene way upstream to the trp operon. It is constitutively on. The trp repressor binds to the operator distal to the promoter blocking the transcription of the operon. All of that is the same as the lac operon. The difference is that the trp repressor is made **inactive by default**. Where the i-protein requires sequestration by lactose so it won't bind the operator, the trp repressor **requires sufficient corepressor** to be turned on so that it can bind the operator. The corepressor is actually tryptophan. The more tryptophan the operon makes, the more trp repressor is activated, and the less tryptophan is made.

The entire system defaults to on. "Make tryptophan" is default-on. "Make tryptophan repressor" is default-on. Just like the i-gene in the lac operon, this tryptophan repressor is somewhere upstream-but-nearby. UNLIKE the i-gene repressor, the trp repressor CAN'T bind the operator by default. It's made, it's there, but until **tryptophan increases in quantity**, the trp repressor can't do anything. So while lactose sequestered the i-gene in the lac operon, the trp repressor is activated by too much tryptophan, thereby limiting excess tryptophan from being made. Because we need more tryptophan to turn off the tryptophan operon, we call **tryptophan the corepressor**.

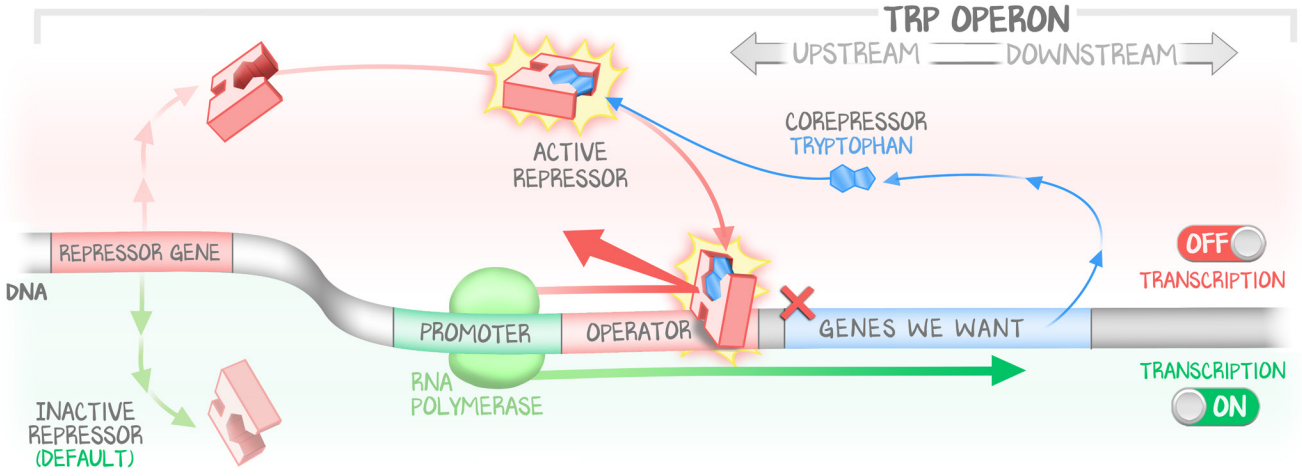

Figure 11.3: Trp Operon Repressor
Trp operon showing the trp-repressor protein distal and upstream to the operon, the absence of the CAP protein, and the expression of the corepressor trp to activate the trp repressor. Identifies the LEAD segment discussed next.

Why the trp operon gets play now is that it has replaced the histidine operon in explaining **attenuation**. The tryptophan repressor isn't enough to stop excess tryptophan from being made. In the lac operon we had a CAP protein enhancer upstream of the promoter region, and the operator suppressor downstream of the promoter. In the trp operon, there is no CAP protein promoter. There is the operator suppressor, but we just said it's not good enough.

Attenuation is possible only in prokaryotes because attenuation requires translation and transcription to happen simultaneously. In eukaryotes, mRNA is packed, exported, and translated in cytoplasm. So because translation and transcription occur simultaneously in prokaryotes, attenuation is possible.

Attenuation is caused by a **leading sequence** that exists after the operator but before the coding segment. Tryptophan is being made. There's too much of it. We need a backup inhibitor because the trp repressor isn't good enough. What we can do instead is use that tryptophan, that excess amino acid, as a speed bump.

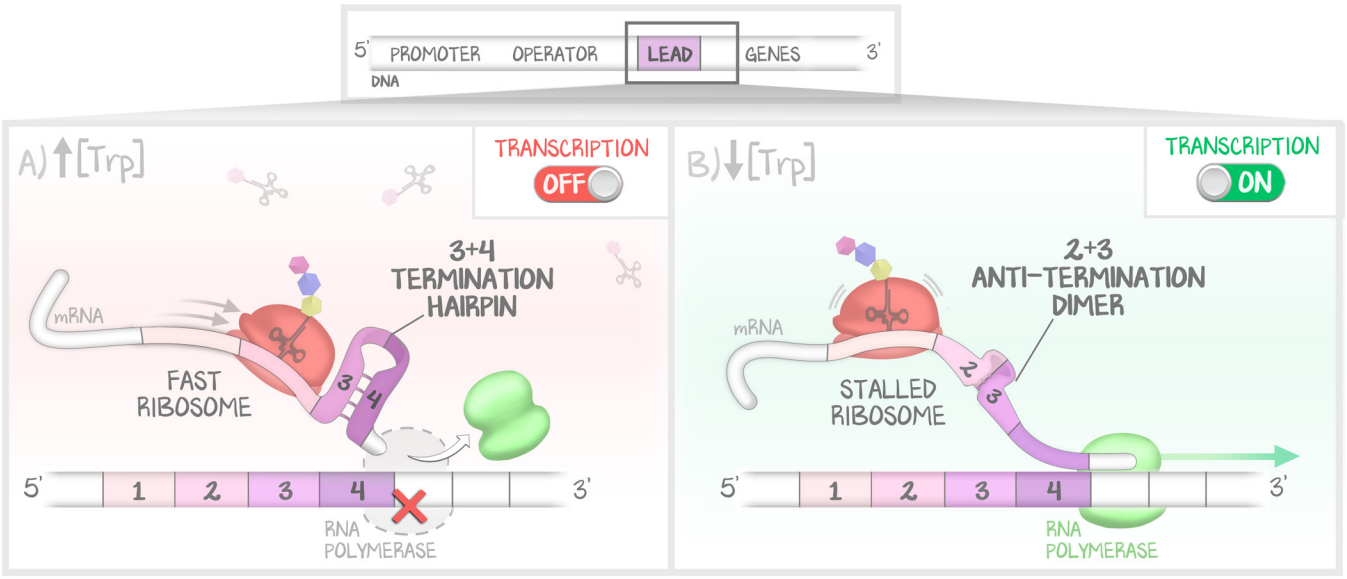

Figure 11.4: Trp Operon Attenuation
Demonstrating the II-III complex as anti-termination signal and III-IV complex as termination.

The RNA polymerase is going down the DNA making mRNA. The ribosome is going down the mRNA making amino acids. There are **4 sequences** in the leading sequence, and each of those can dimerize with its neighbor if and only if there isn't a protein on it. That means if there is a ribosome on the mRNA strand, the mRNA cannot dimerize with that sequence. If the RNA polymerase hasn't finished that sequence, then the unfinished sequence cannot dimerize. When two adjacent sequences are free, they dimerize. That dimerization will have a different effect depending on which sequences pair.

The first sequence has a bunch of UGG's (codon for tryptophan) which act to rate-control the velocity of the ribosome. If there are a lot of tryptophans, the ribosome can easily find the tryptophan tRNA, builds the amino acid sequence, and passes through sequence 1 quickly. If there are few tryptophans, the ribosome has trouble finding tryptophan tRNAs, and lingers, passing through sequence 1 slowly.

RNA polymerase is going. Ribosomes are slow. mRNA sequences can dimerize with each other. If tryptophan is low, the ribosome is slower, and the RNA polymerase is fast. This means the ribosome will get stuck on sequence 1 waiting for the tryptophan, and RNA polymerase will make sequences 2 and 3. While the ribosome is still on 1, and the RNA polymerase moves into 4, 2 and 3 are liberated. Having nothing on them, they dimerize. This gives an **anti-termination signal** (HURRY UP, MAKE MORE TRYPTOPHAN!).

If, on the other hand, there is a lot of tryptophan, so that when the ribosome gets to sequence 1 it moves past it no problem, it then moves onto sequence 2. Well, sequences 1 and 2 can't dimerize with each other: there is a ribosome sitting on 2. Sequences 2 and 3 can't dimerize, either: the ribosome is on sequence 2. The RNA polymerase is faster, so as the RNA polymerase leaves sequence 4 to begin coding the genes of the operon, it exposes sequences 3 and 4. Sequence 3 and sequence 4 of the **mRNA strand** dimerize. When they do, what results is a GC-rich hairpin turn with a series of UUUUUU following. Sound familiar? Yep, that's the stop signal for transcription. The entire operon is terminated. There is a balance of the formation of the GC-hairpin-UUUUUU stopping operon formation (trp high) and the 2-3 pairing which stimulates the system (trp low). But what you need to see is that structural **changes in the mRNA cause gene expression to vary simply by varying the concentration of the operon product**.

Conclusion

Not worth the squeeze. This is 270 stuff. Skip this if it sounds like a lot. It is. For very little in return. Also realize this is prokaryotic regulation. I doubt you'll see too many prokaryotes in your bustling practice.